基于异构数据的联邦学习自适应差分隐私方法研究

徐茹枝, 仝雨蒙✉,, 戴理朋

华北电力大学控制与计算机工程学院，北京 102206

Research on Federated Learning Adaptive Differential Privacy Method Based on Heterogeneous Data

XU Ruzhi, TONG Yumeng✉,, DAI Lipeng

School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China

通讯作者: 仝雨蒙 tongym02@163.com

收稿日期: 2024-09-28

基金资助: 国家重点研发计划(62372173)

Received: 2024-09-28

作者简介 About authors

徐茹枝（1966—），女，江西，教授，博士，主要研究方向为 AI 安全、智能电网 。

仝雨蒙（2002—），女，河北，硕士研究生，主要研究方向为联邦学习 E-mail：tongym02@163.com。

戴理朋（1999—），男，安徽，硕士研究生，主要研究方向为联邦学习、差分隐私 。

摘要

在联邦学习中，由于需要大量的参数交换，可能会引发来自不可信参与设备的安全威胁。为了保护训练数据和模型参数，必须采用有效的隐私保护措施。鉴于异构数据的不均衡特性，文章提出一种自适应性差分隐私方法来保护基于异构数据的联邦学习的安全性。首先为不同的客户端设置不同的初始隐私预算，对局部模型的梯度参数添加高斯噪声；其次在训练过程中根据每一轮迭代的损失函数值，动态调整各个客户端的隐私预算，加快收敛速度；接着设定一个可信的中央节点，对不同客

户端的局部模型的每一层参数进行随机交换，然后将混淆过后的局部模型参数上传到中央服务器进行聚合；最后中央服务器聚合可信中央节点上传的混淆参数，根据预先设定的全局隐私预算阈值，对全局模型添加合适的噪声，进行隐私修正，实现服务器层面的隐私保护。实验结果表明，在相同的异构数据条件下，相对于普通的差分隐私方法，该方法具有更快的收敛速度以及更好的模型性能。

关键词： 联邦学习; 异构数据; 差分隐私; 高斯噪声

Abstract

In federated learning, the need for a large amount of parameter exchange may lead to security threats from untrusted participating devices. In order to protect training data and model parameters, effective privacy protection measures must be taken. Given the imbalanced nature of heterogeneous data, this paper proposed an adaptive differential privacy method to protect the security of federated learning based on heterogeneous data. Firstly, different initial privacy budgets were set for different clients, and Gaussian noise was added to the gradient parameters of the local model; Secondly, during the training process, the privacy budget of each client was dynamically adjusted based on the loss function value of each iteration to accelerate convergence speed; Then, set a trusted central node to randomly exchange the parameters of each layer of local models from different clients, and then uploaded the confused local model parameters to the central server for aggregation; Finally, the central server aggregated the obfuscation parameters uploaded by trusted central nodes, added appropriate noise to the global model based on a pre-set global privacy budget threshold, and performed privacy correction to achieve server level privacy protection. The experimental results show that under the same heterogeneous data conditions, compared to ordinary differential privacy methods, the adaptive differential privacy method proposed in this paper has faster convergence speed and better model performance.

Keywords： federated learning; heterogeneous data; differential privacy; Gaussian noise

徐茹枝, 仝雨蒙, 戴理朋. 基于异构数据的联邦学习自适应差分隐私方法研究[J]. 信息网络安全, 2025, 25(1): 63-77
doi:10.3969/j.issn.1671-1122.2025.01.006

XU Ruzhi, TONG Yumeng, DAI Lipeng. Research on Federated Learning Adaptive Differential Privacy Method Based on Heterogeneous Data[J]. Netinfo Security, 2025, 25(1): 63-77 doi:10.3969/j.issn.1671-1122.2025.01.006

0 引言

联邦学习是一种解决数据隐私和安全问题的机器学习技术，可以将原始数据保留在各参与方设备本地，各参与方协同训练一个共有模型，不需要从各参与方收集需要的数据，这不仅能保证用户的数据隐私和安全，而且能够联合各参与方的优质隐私数据，提高模型的准确性和通用性[1]。然而，在实际应用中，联邦学习面临参与者之间的数据异构性问题，给联邦学习的隐私保护带来了诸多挑战[2]。

异构数据主要指来自不同源、格式或特征的数据。数据异构性严重影响了联邦学习的训练效率和精度，降低了模型性能。LI[3]的研究表明，由于不同设备或用户之间的数据分布差异，传统的联邦学习算法在处理异构数据时难以达到理想的训练效果；ZHAO[4]实验验证了在数据分布差异较大时，传统的联邦学习算法难以收敛到全局最优解，降低了训练效率和模型性能。在大数据时代下，数据增速高、种类多且传播快，更容易被恶意攻击者获取并传播，隐私泄露成为当今社会发展的隐患。因此，如何处理异构数据来保证联邦学习的效果，同时保证模型训练过程中的数据隐私性，是当前联邦学习研究领域亟需解决的问题之一。

目前，联邦学习通过差分隐私[5]、安全多方计算[6]、同态加密、基于可信的硬件[7]等方法来增强数据的隐私性和安全性。然而，在基于异构数据的联邦学习训练过程中，传统的隐私保护方法往往忽视不同客户端数据间的异构性。MOHAMMADI[8]等人通过引入隐私增强的 FedPaq 算法，探究了梯度扰动参数对联邦学习模型性能的影响，在保证通信效率的同时保证了异构数据的隐私性。GONG[9]等人利用未标记的公共数据进行单向离线知识蒸馏来保护本地数据隐私，并引入新颖的注意力约束来提取局部模型的知识，在局部共识和多样性之间达到了平衡，解决了数据异质性问题。GAO[10]等人提出一种名为异构联邦迁移学习的方法，使用同态加密和安全共享技术解决了异构数据场景下的隐私保护问题。NOBLE[11]等人结合差分隐私（DP）理论和 SCAFFOLD 优化算法来解决异构数据条件下的联邦学习隐私保护问题，即保证了隐私又提高了模型训练效率和准确性。根据当前研究现状，在异构数据场景下，若对所有客户端采取统一的隐私预算分配，可能会导致算法性能的显著下降。鉴于不同数

据端的数据量不平衡及其对隐私预算的需求，实施个性化的隐私保护策略变得至关重要。通过为不同客户端以及不同的训练阶段提供个性化隐私保护方法，可以更有效地保障模型参数的隐私安全，同时保证算法性能不受较大影响。

综上所述，多端数据的异构性已逐渐凸显为制约联邦学习进一步发展的关键因素。因此，本文对于联邦学习在异构数据条件下的隐私保护研究具有重要的意义，主要研究内容包括以下 3 个方面。

1） 使用差分隐私方法对局部模型进行自适应加噪，对联邦学习的局部模型参数添加高斯噪声，并动态调整各个客户端的隐私预算，保证了隐私保护的有效性。

2）利用梯度参数洗牌方法，由一个可信的中央节点重新排列各个客户端上传的模型参数，提高了隐私保护的强度。

3）设置一个全局隐私预算阈值，由中央服务器控制添加噪声的大小，调整全局隐私水平，实现了服务器层面的隐私保护。实验结果表明，针对异构数据的隐私保护问题，本文提出的自适应差分隐私方法具有更优良的模型性能、更低的通信成本和更快的收敛速度。

1 联邦学习

联邦学习是一种分布式机器学习框架，它允许参与者在不共享本地隐私数据的情况下，通过中央服务器协作训练出共享的全局模型。其核心思想是各参与者训练一个模型，然后在各自模型上进行数据交流，最终中央服务器进行模型聚合，得到全局模型。各客户端之间交换的数据无法被猜测到，这种方法有效避免了数据传输中的隐私泄露问题[12]。

1.1 联邦学习概述

假设联邦学习系统有 n 个客户端$\{u_1,\cdots,u_n\}$，各客户端持有它们的私有数据$\{D_1,\cdots,D_n\}$，且各客户端不能直接访问其他客户端的数据。在传统的机器学习中，由服务器收集 n 个客户端的本地数据进行统一训练得到全局模型 M_{global}，该训练过程会

将服务器的本地数据暴露给服务器。在联邦学习中，由客户端对本地数据进行模型训练得到本地模型，然后服务器收集各客户端上传的本地模型并聚合得到全局模型 M_{fed_global}Mfed_global。

典型的联邦学习系统架构如图 1 所示，联邦学习系统通过从分布式设备中收集训练信息来学习一个模型，在中央服务器的协调下，客户端（手机、电脑、平板等设备）协同训练模型，经过多轮交互，模型不断优化直至收敛，完成训练。具体实现过程如下[13]。

图 1

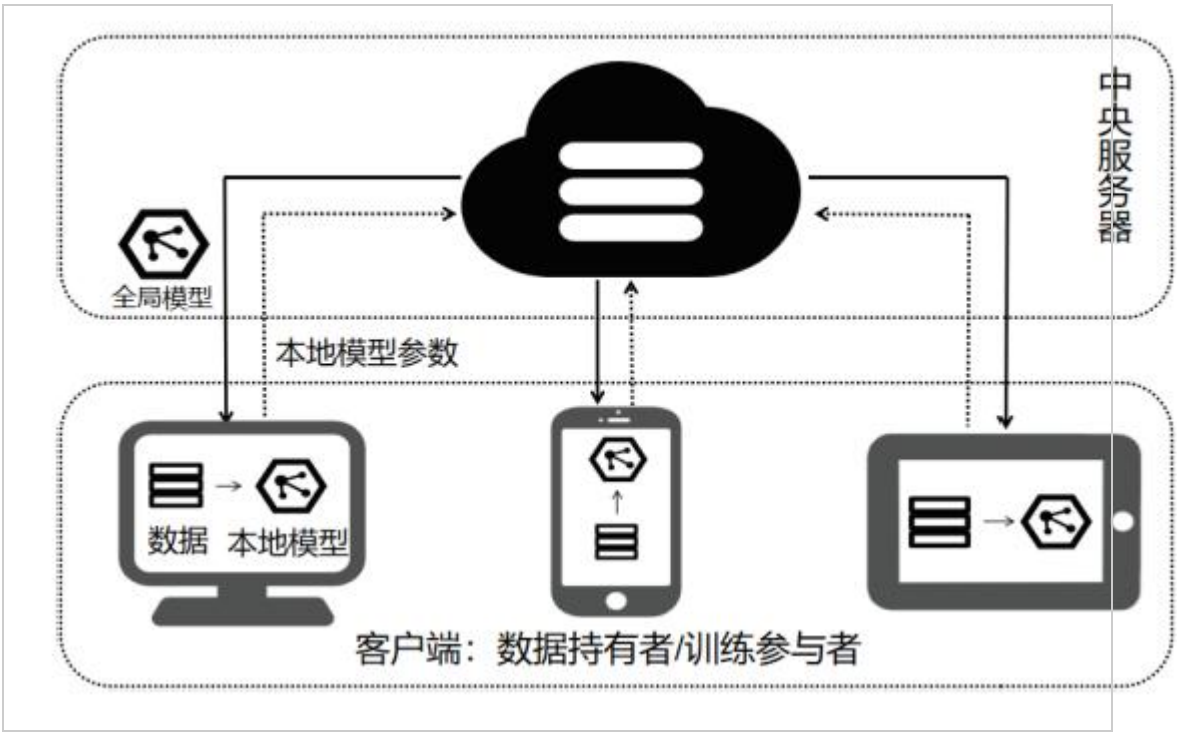


图 1　联邦学习系统架构

1）服务器选择参与本轮训练的客户端设备，然后由服务器建立初始模型，将初始模型的结构与参数发送给参与训练的各客户端。

2）各参与方从服务器获取全局模型作为本地初始模型，根据本地数据进行模型训练，完成训练后将更新的模型参数上传到服务器。

3）服务器对收到的各参与方上传的模型参数进行聚合得到全局模型，然后将该全局模型发送给下一轮参与训练的客户端，各参与方更新本地模型后进行下一轮模型训练。

4）联邦学习系统重复步骤 2）和步骤 3）训练过程，直至模型达到收敛条件，服务器将最终的训练模型发送给各客户端。

联邦平均算法（FedAvg）是一种常用的联邦学习算法，通过加权平均来聚合模型参数，假设参与模型训练的客户端总数为 KK，公式（1）和公式（2）为 FedAvg 算法定义的目标函数。

$$min_{\omega \in R^d} f(\omega)=\sum_{k=1}^{K} \frac{n_k}{n} F_k(\omega) \quad \min\omega \in R\, df(\omega)=\sum k=1Knknn Fk(\omega)$$

（1）

$$F_k(\omega)=\frac{1}{n_k}\sum_{i \in P_k} f_i(s_i;\omega_i) \quad Fk(\omega)=1nk\sum i\in Pkfi(si;\omega i)$$

（2）

其中，$F_k(\omega)$Fk(ω)表示第 kk 个客户端的本地损失函数，P_kPk 表示第 kk 个客户端的训练样本，$n_k=\left|P_k\right|$nk=|Pk|表示样本个数，$s_i$si 表示第 ii 个样本个体，ωω表示神经元的权重矩阵，ω_iωi 表示模型在 $s_i$si 训练得到的权重矩阵，$f_i(s_i,\omega_i)$fi(si,ωi)表示模型在 $s_i$si 的损失函数。

一次迭代即局部模型的一次更新，用 bb 表示一个 batch。联邦学习在第 kk 个客户端上的局部模型迭代过程ω_kωk 如公式（3）所示。

$$\omega_k \leftarrow \omega_k-\eta\frac{1}{|b|}\sum_{i \in b}\nabla f(s_i;\omega_i) \quad \omega k\leftarrow \omega k-\eta|b|\sum i\in b\nabla f(si;\omega i)$$

（3）

全局模型通过对各个局部模型进行加权求和得到，令ω_{k}^{t+1}表示第 t 轮训练结束后客户端 k 得到的局部模型，模型的聚合如公式（4）所示。

$$\omega^{t+1}=\sum_{k\in S_t} n_k \omega_k^{t+1} \quad (4)$$

（4）

算法 1 描述了 FedAvg 算法的聚合过程。

算法 1 联邦平均算法（FedAvg）

输入：参与训练的客户端总数 K，客户端本地迭代次数 E，初始化全局模型ω_0

输出：训练完成的全局模型ω_{T-1}

1.for t = 0,…,T-1 do

2. 服务器随机选择 K 个客户端的一个子集 S_t/*每个设备 k 被选择的概率为 p_k*/

3. 服务器发送当前模型参数ω_t 至所有被选择的客户端

4. 每个客户端设备 $k\in S_t$ 在本地进行 E 次随机梯度下降迭代，步长为η，更新本地模型为ω_{k}^{t+1}

5. 每个客户端设备 $k \in S_t$k∈St 上传本地更新模型 ω_{kt+1}ωt+1k 至服务器

6. 服务器对客户端上传的模型参数进行聚合：

$\omega_{t+1}=\omega t+1=1K\sum k\in S_t\omega_{kt+1}1K\sum k\in S t\omega t+1k$

7.end for

8.return $\omega_{T-1}\omega T-1$

1.2 联邦学习中的数据异构分类

异构数据场景是联邦学习面临的一个难题。具体而言，由于各客户端参与训练的数据虽然各自独立分布，但却不遵循相同的采样方法即数据非独立同分布（Non-Identically and Independently Distributed，Non-IID）问题，这一问题导致模型精度大幅下降，为传统的联邦学习算法带来了极大的困难。

在联邦学习中，考虑一个特征为 *x* 标签为 *y* 的监督任务，联邦学习的数据统计涉及两个方面的采样：首先在所有可用设备 QQ 中抽样出一个设备 ii 用于访问数据样本，其次从该客户端的本地数据分布中抽取一个样本数据 $P_i(x,y)$Pi(x,y)。对于不同的客户端设备 ii 和 jj，其样本数据 $P_j(x,y)\neq P_i(x,y)$Pj(x,y)≠Pi(x,y)，将 $P_i(x,y)$Pi(x,y) 表示为 $P_i(y|x)P_i(x)$Pi(y|x)Pi(x)和 $P_i(x|y)P_i(y)$Pi(x|y)Pi(y)可以更精确地表征联邦学习中不同异构数据的差异，根据 YANG[14]等人的研究，联邦学习的数据异构类型有以下 5 种。

1）特征分布偏移

设备之间的 $P(y|x)$P(y|x) 是共享的，即对于不同客户端设备 ii 和 jj，有 $P_i(y|x)=P_j(y|x)$Pi(y|x)=Pj(y|x)，但特征的边际分布 $P_i(x)$Pi(x)可能因客户端的不同而有所差异。例如，在语音识别领域，不同用户或设备录制的语音数据，其音频质量、背景噪音、语速、语调等特征都可能存在显著差异。

2）标签分布偏移

设备之间的 $P(x|y)$P(x|y) 是共享的，即对于不同客户端设备 ii 和 jj，有 $P_i(x|y)=P_j(x|y)$Pi(x|y)=Pj(x|y)，但标签的边际分布 $P_i(y)$Pi(y)可能因客户端的不同而有所差异。例如，在医疗健康领域，不同医院或医疗机构可能收集到不同疾病类型的患者数据，由于不同医院的专业领域、地理位置或患者群体差异，某些疾病的标签分布可能在这些机构之间呈现出显著的不均衡。

3）标签相同、特征不同

即使 $P(y)$P(y)相同，条件分布 $P_i(x|y)$Pi(x|y)也可能因客户端的不同而有所差异。即对于不同的客户端，相同的标签 *y* 可能具有不同的特征 *x*。例如，英国的短毛猫和暹罗猫虽然都属于“猫”这一标签，但它们的外观特征却大相径庭。

4）特征相同、标签不同

即使 $P(x)$P(x)相同，条件分布 $P_i(y|x)$Pi(y|x)也可能因客户端不同而有所差异。由于个体偏好和认知差异，同一组训练数据中相同的特征向量可能具有不同的标签。例如，不同鉴赏者对同一画作持有不同的评价。

5）数量倾斜或不平衡

不同的客户端可以持有数量差异很大的数据。例如，在某些情况下，特定标签的样本数量可能在客户端之间分布不均，这种数量倾斜或不平衡的现象会对模型的预测性能产生不良影响。

由于数据的异构性，每个客户端的本地数据集分布与全局分布存在显著差异。这种差异导致各方的局部模型最优方向与全局模型最优方向不一致，进而产生了局部更新偏移的现象。在局部训练阶段，每个模型基于其本地数据，会朝着各自的局部最优方向进行更新。然而，这些局部最优点可能远离全局最优点，从而影响全局模型的收敛效果。因此，相较于非异构数据设置下的联邦学习，这种情况下的全局模型精度会显著降低。图 2 是 Non-IID 设置下的模型偏移示例，表示的是基于异构数据下的 FedAvg 问题[15]。如图 2 左侧所示，在非异构数据设置下，全局最优 W* w* 接近于客户端的局部最优 W* 1w1* 和 W* 2w2* ，因此对当前轮次客户端上传的模型参数 Wt1w1t 和 Wt2w2t 进行平均聚合后的模型 Wt+1wt+1 也接近全局最优 W* w* 。如图 2 右侧所示，在异构数据设置下，W* w* 远离 W* 1w1* ，所以聚合模型 Wt+1wt+1 可能会远离全局最优 W* w* 。

图 2

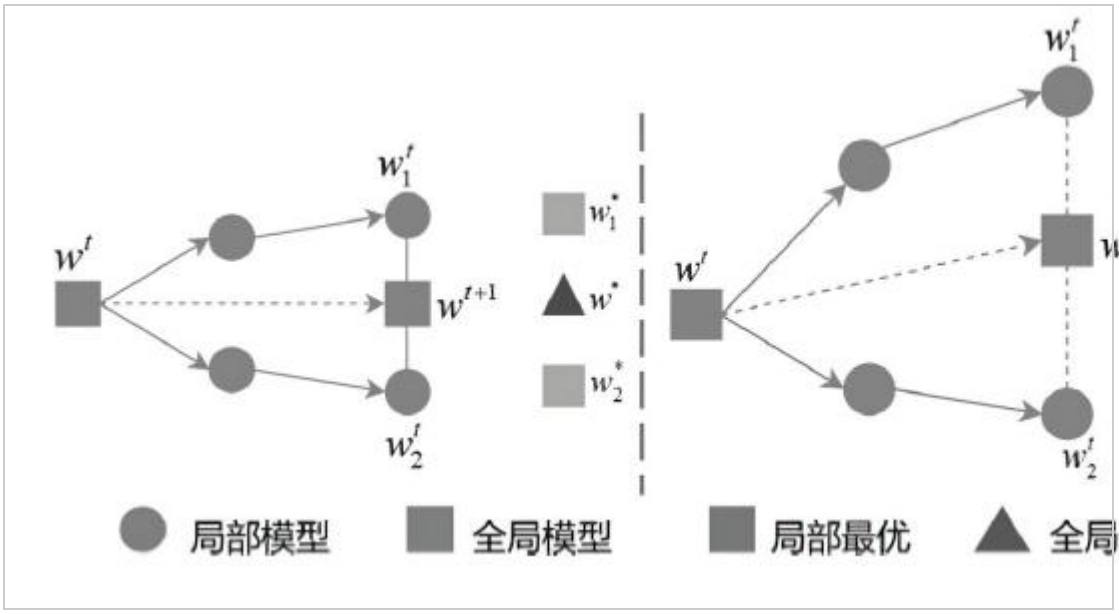


图 2　Non-IID 设置下的模型偏移示例

本文重点关注特征分布偏移的数据分布场景下的联邦学习隐私保护研究。

1.3 差分隐私定理

差分隐私（Differential Privacy）是一种在数据分析和机器学习领域广泛应用的隐私保护技术，由 DWORK[16]在 2006 年提出，其核心思想是通过在数据处理中引入随机性来确保个体信息的隐私性，使得攻击者无法准确推断出特定个体的信息。在联邦学习中，差分隐私通过添加噪声来实现，确保算法输出不受数据源差异或存在与否的影响，从而保护参与方的个体数据不被反推。噪声添加通常依赖于拉普拉斯分布机制（Laplace Mechanism）或高斯分布机制（Gaussian Mechanism）等，并在满足隐私预算（Privacy Budget）的前提下，力求最小化对数据分析结果的影响。

(ε,δ)-(ε,δ)-差分隐私可以为机器学习算法 M 的数据集提供隐私保护，其形式化定义如公式（5）所示。

$$Pr[M(D_i)\in S]\leqslant e^{\varepsilon}Pr[M(D_i')\in S]+\delta \Pr[M(Di)\in S]\leqslant e^{\varepsilon}\Pr[M(Di')\in S]+\delta$$

（5）

其中，S 表示算法 M 输出的任意一个子集，$\varepsilon>0$ 称为隐私预算或者隐私保护强度，表示相邻两个数据集之间差异的上限。ε越小，隐私保护强度越大，需要添加的扰动噪声也越大；反之，ε越大，隐私保护强度越小。δ是一个非 0 的实数，表示松弛程度，即满足上述不等式的概率。

$\varepsilon-$差分隐私原本是一种严格的隐私保护技术，不包括δ项，但公式（5）中提出的松弛差分隐私则允许一定程度的松弛，即允许算法以特定概率δ被打破，从而提供了更灵活的隐私保护方案。这种松弛机制在一定程度上放松了差分隐私的严格定义，使其能够更好地适应不同场景的需求。差分隐私不仅具有出色的通用性和可扩展性，可以广泛应用于各种数据分析和利用场景，而且它还可以与其他隐私保护方法相结合，进一步提升隐私保护效果。徐茹枝[17]等人通过基于联邦学习的中心化差分隐私保护算法 DP-FEDAC，解决了联邦学习中的半诚实和恶意客户端对全局模型的差分攻击造成的隐私泄露问题。

2 自适应差分隐私方法

针对联邦学习训练过程中异构数据的隐私保护，本文提出一种联邦学习自适应差分隐私方法，如图 3 所示。该方法主要包括局部模型自适应性加噪、模型参数洗牌以及全局模型隐私保护修正 3 个部分。

图 3

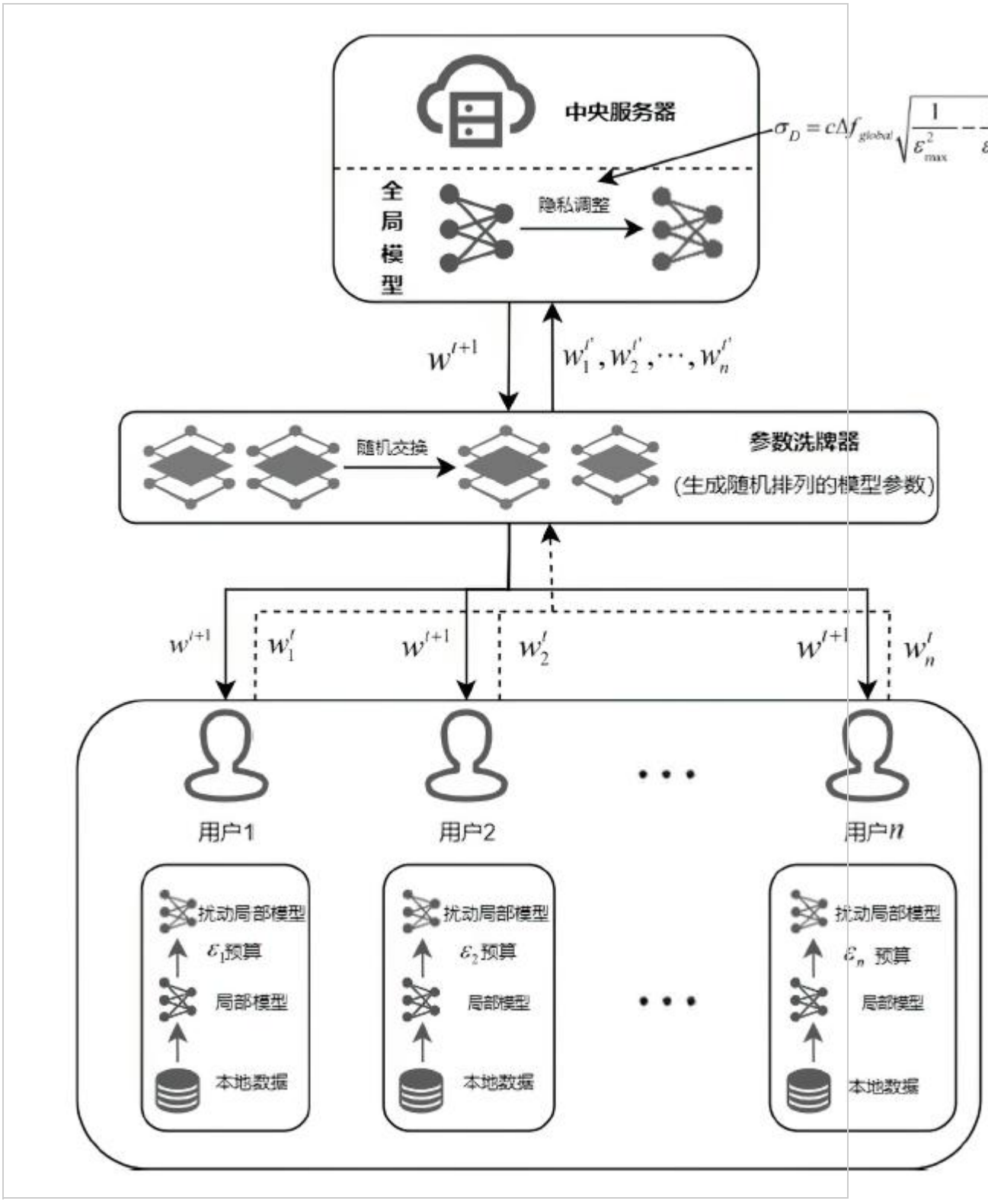


图 3　自适应差分隐私方法框架

2.1 局部模型自适应加噪

本文方法的局部模型自适应加噪部分通过差分隐私方法对局部模型参数添加扰动实现对数据的隐私保护，主要包括设定初始隐私预算、调整隐私预算、梯度裁剪和添加噪声扰动。

2.1.1 设定初始隐私预算

在异构数据分布环境下，不同客户端的隐私保护需求和数据质量各不相同，在这种情况下，如果所有的客户端设定相同的隐私预算，就无法满足不同客户端的个性化隐私保护需求，并且持有较低数据质量的客户端训练的模型效果会受到更大的噪声影响。所以，对各客户端设定不同的初始隐私预算来添加高斯噪声，即对于客户端集合$\{C_1,C_2,\cdots,C_n\}$分别设定符合需求的初始隐私预算$\{\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n\}$。

2.1.2 调整隐私预算

各客户端在进行新一轮模型训练前，根据前一轮训练结果的损失函数（*loss*）值动态调整客户端的隐私预算。在模型训练的初始阶段，*loss* 值较大，隐私预算较低，此时对模型的梯度参数添加较大的高斯噪声，完成数据的隐私保护。随着模型训练的进行，*loss* 值减小，隐私预算逐步增加，模型的训练精度提高，此时减小对模型的梯度参数添加的高斯噪声，有利于联邦学习模型快速收敛。

本小节提出了线性调整和指数调整两种调整预算的方法，如公式（6）和公式（7）所示。

$$\varepsilon_{ji}=\varepsilon_i+k(loss_{Maxi}-loss)\varepsilon ij=\varepsilon i+k(lossiMax-loss) \tag{6}$$

$$\varepsilon_{ji}=\varepsilon_i\times exp(k(loss_{Maxi}-loss))\varepsilon ij=\varepsilon i\times exp\,f_O\{k(lossiMax-loss)) \tag{7}$$

公式（6）和（7）中，*i* 表示客户端 *i*，*j* 表示第 *j* 轮联邦学习训练，*k* 表示调整系数，该调整系数根据模型训练效果确定，$loss_{Maxi}$lossiMax 表示客户端 *i* 的最大 *loss* 值，本小节将各客户端第一轮模型训练得到的 *loss* 值定义为该客户端的最大 *loss* 值。

线性调整方法简单直观，计算复杂度较低，具有更高的透明度和可控性，便于理解，易于实现；指数调整方法灵活性和敏感性较高，根据 *loss* 值的变化非线性方式调整隐私预算，保证模型性能的同时更加细致地控制了隐私保护的效果。这两种调整方法的目的都是随着 *loss* 值的降低增加隐私预算。

在实际场景中，应该基于具体的数据特征、隐私预算需求来确定选择哪种隐私预算调整方法，本文在实验部分会分析两种调整方法在异构数据场景下对模型训练的影响。

2.1.3 梯度裁剪

根据差分隐私公式可知，差分隐私添加的噪声大小与函数的敏感度高度相关，函数的敏感度由梯度范数决定，因此，为了达到保护数据隐私的同时提高模型的训练精度和性能并且降低计算的复杂度，利用梯度裁剪方法来控制噪声的大小。

常见的梯度裁剪方法分为基于阈值直接裁剪梯度和按范数裁剪梯度。基于阈值直接裁剪梯度方法是在给定的范围内约束梯度，当梯度值超过该范围时就会被直接裁剪为边界值，这种方法是最简单直接的梯度裁剪方法。因为函数的敏感度由梯度的范数决定，所以本文选择基于 L2 范数裁剪梯度的方法，具体裁剪方法如下。

假设给定的裁剪最大范数 $Max_norm = 2.0$Max_norm = 2.0，假设给定参数梯度 $grad=[1.5,2.5,2.0]$grad=[1.5,2.5,2.0]，计算该参数梯度的 L2 范数 $grad_2=1.5_2+2.5_2+2_2-------------\sqrt{}\approx3.5355$grad2=1.52+2.52+22≈3.5355，计算缩放系数

$clip_coef=Max_normgrad_2=23.5355\approx0.5657$clip_coef=Max_normgrad2

=23.5355≈0.5657。

裁剪结果

$grad_cliped=clip_coef\times grad=[0.8486,1.4143,1.1314]$grad_

cliped=clip_coef×grad=[0.8486,1.4143,1.1314]。

2.1.4 添加噪声扰动

高斯噪声，也称为正态分布噪声，是指它的概率密度函数服从高斯分布（即正态分布）$N(0,\sigma2)$N(0,σ2)的一类噪声，均值为 0，标准差$\sigma\sigma$为常数，其概率密度函数定义如公式（8）所示。

$$f(x)=12\ \pi\ \sqrt{}\sigma exp(-(x-\mu)_22\sigma_2)$$f(x)=12 π σexp[0](−(x−μ)22σ2)

（8）

其中，xx 为随机变量，$\mu\mu$为均值，$\sigma\sigma$为标准差。如图 4 所示，可以看到$\sigma\sigma$的大小决定高斯分布的幅度。

图 4

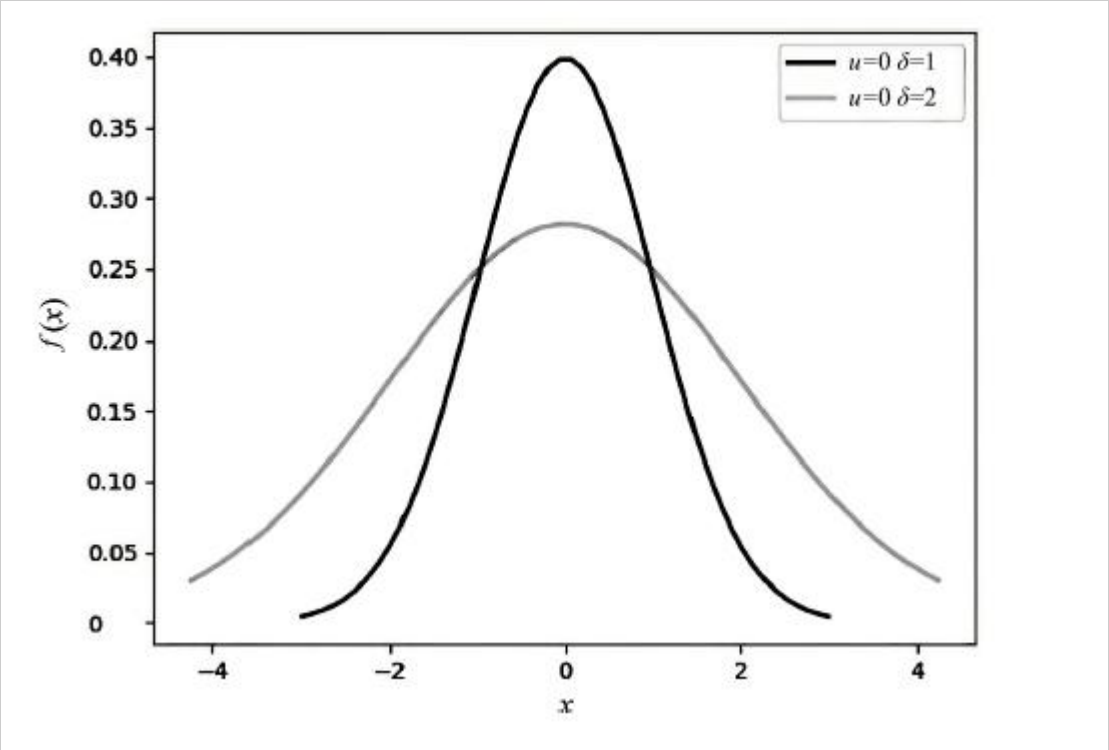


图 4　高斯分布

在差分隐私中，标准的差分隐私太过严格，高斯机制通过向查询结果 $f(D)$f(D)中添加高斯噪声 yy 来实现松弛的差分隐私，如公式（9）所示。

$$M(D)=f(D)+y$$M(D)=f(D)+y

（9）

查询函数 ff 的敏感度衡量了相邻数据集 DD 和 D'D′的查询结果的最大差异，差分隐私通过 L2 范数计算敏感度并根据该结果生成符合高斯分布的噪声，如公式（10）所示。

$$\Delta f=max_{D,D'}\parallel f(D)-f(D')\parallel_2$$Δf=maxD,D′‖f(D)−f(D′)‖2

（10）

在联邦学习模型中，客户端 *i* 在第 *k* 轮迭代过程中，可能会由于噪声扰动过大导致梯度爆炸问题，使得模型最终无法收敛得到全局最优解，因此使用 L2 范数对客户端 *i* 的训练结果进行梯度裁剪。设定裁剪阈值为 MM，当 $//gki//2\leq M$‖gik‖2≤M 时，不执行梯度裁剪，当 $//gki//2>M$‖gik‖2>M 时，对梯度进行裁剪操作，$gki=gki\cdot M/gki/2$gik=gik· M‖gik‖2。

对完成裁剪后的梯度添加高斯噪声 $N(0,\Delta f2\sigma 2)$N(0,Δf2σ2)，如公式（11）所示。

$$gki\sim\leftarrow gki+N(0,\Delta f2\sigma 2)$$gik~ ←gik+N(0,Δf2σ2)

（11）

根据公式（5）描述的松弛差分隐私定理，对于任意 $\epsilon\in(0,1),\delta\geq e-(\sigma\epsilon)2/21.25$ϵ∈(0,1),δ≥e−(σϵ)2/21.25，噪声 $Y\sim N(0,\Delta f2\sigma 2)$Y~ N(0,Δf2σ2)满足 $(\varepsilon,\delta)-DP$(ε,δ)−DP，松弛差分隐私的满足条件如公式（12）所示。

$$Pr[M(d)\in S]\leq e\epsilon Pr[M(d')\in S]+\delta$$Pr[M(d)∈S]≤eϵPr[M(d′)∈S]+δ

（12）

其中，MM 表示差分隐私保护算法，σσ表示高斯分布标准差，εε表示隐私预算，与噪声大小负相关，ΔfΔf 表示算法的敏感度，与噪声大小正相关，δδ表示可以容忍的违反严格差分隐私的概率值。

将客户端添加噪声后的梯度值上传到服务器，得到添加噪声后的梯度更新之和如公式（13）所示。

$$\Delta g_{sum} \leftarrow \sum_{i=1}^{n}(g_{ki}+N(0,\Delta f^2\sigma^2))\Delta gsum \leftarrow \sum i=1n(gik+N(0,\Delta f2\sigma 2)) \tag{13}$$

将梯度更新之和除以参与模型训练的客户端数量 n，与 t 时刻的全局模型参数相加，得到的新一轮的全局模型参数如公式（14）所示。

$$w_{t+1} \leftarrow w_{t+1}{}_n\sum_{i=1}^{n}(g_{ki}+N(0,\Delta f^2\sigma^2))wt+1 \leftarrow wt+1n\sum i=1n(gik+N(0,\Delta f2\sigma 2)) \tag{14}$$

实现局部模型自适应加噪的算法伪代码如算法 2 所示。

算法 2 局部模型自适应加噪算法

输入：客户端隐私预算$\{\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n\}\{\varepsilon 1,\varepsilon 2,\cdots,\varepsilon n\}$，全局模型参数和客户端上传梯度$(w_t,g_{ki})(wt,gik)$，梯度裁剪阈值 MM

输出：添加高斯扰动后的全局模型参数

客户端：

1. 计算客户端 j 当前轮次的隐私预算（线性调整）：$\varepsilon_{ji}=\varepsilon ij=\varepsilon_i+k(loss_{Maxi}-loss)\varepsilon i+k(lossiMax-loss)$

2. 根据算法敏感度计算高斯噪声分布 $N(0,\Delta f^2\sigma^2)$N(0,Δf2σ2)

3. 根据裁剪阈值对局部模型梯度进行裁剪：$g_{ki}=g_{ki}\cdot M/\max(\|g_{ki}\|_2)$gik=gik⋅ M/max(‖gik‖2)

4. 对梯度添加高斯扰动：$g_{ki}\sim \leftarrow g_{ki}+N(0,\Delta f^2\sigma^2)$gik~ ←gik+N(0,Δf2σ2)服务器：

5. 梯度聚合：

$\Delta g_{sum}\leftarrow\sum_{i=1}^{n}(g_{ki}+N(0,\Delta f^2\sigma^2))$Δgsum←∑i=1n(gik+N(0,Δf2σ2))

6. 计算新一轮全局模型：

$w_{t+1}\leftarrow w_{t+1}\frac{n}{\sum_{i=1}^{n}(g_{ki}+N(0,\Delta f^2\sigma^2))}$wt+1←wt+1n∑i=1n(gik+N(0,Δf2σ2))

)

7. return w_{t+1}wt+1

2.2 梯度参数洗牌

在算法 2 的各客户端上传梯度和中央服务器进行梯度聚合之间，设置了一个可信的中央节点，对各客户端上传的梯度添加扰动后再上传到中央服务器进行聚合操作。

GIRGIS[18]等人提出一种联邦学习的梯度参数洗牌方法，通过一个参数洗牌器将客户端上传的压缩梯度进行随机排列之后再上传到中央服务器，切断了服务器与客户端之间的直接联系，有效地降低了推理攻击的可能性，但是攻击者仍然有可能通过梯度相似性判断出梯度与客户端之间的联系。

本节提出的梯度参数洗牌方式，与文献[18]的相同之处在于本节通过设定一个可信的中央节点，即梯度参数洗牌器，重新排列梯度，排列方法为遍历梯度的各层参数，通过随机数方法随机交换各客户端

上传梯度的相应层的参数。如图 5 所示，遍历第 l 层梯度参数时，随机选择两个客户端 m、n，然后交换 m、n 的第 l 层梯度参数，在当前层可以进行多轮交换来达到更好的参数洗牌效果。

图 5

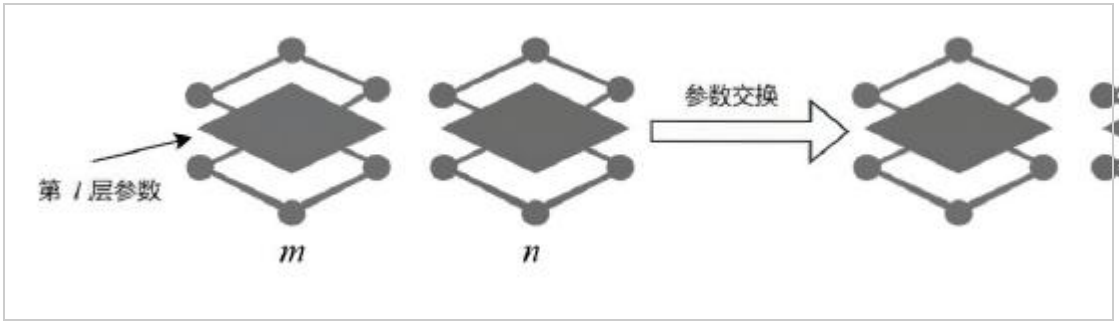


图 5　第 l 层参数交换

这种参数洗牌方法基本解除了交换后的梯度参数和相应客户端的关联，降低了恶意服务器进行推理攻击的可能性，并且没有影响中央服务器对梯度参数的求和平均操作。

2.3 全局模型隐私保护修正

在 2.1 节中，客户端根据用户的隐私偏好来添加合适的噪声，虽然能够实现个性化差分隐私保护，但是这种方式会降低联邦学习模型的训练效率，为了解决这一问题，本节提出设置一个全局差分隐私的隐私预算阈值，在中央服务器端控制添加噪声的大小。

首先，设置一个全局隐私预算阈值ε_{max}εmax。然后，中央服务器端进行参数聚合时，计算当前的全局隐私预算，如公式（15）所示。

$$\varepsilon=\sum_{i=1}^{n}\varepsilon_i'\varepsilon=\sum_{i=1}^{n}\varepsilon_i'$$

（15）

如果参数聚合结果$\varepsilon<\varepsilon_{max}$ε<εmax，中央服务器直接将模型参数下发给客户端以保证训练的模型质量；否则，添加满足全局隐私预算阈值ε_{max}εmax 的噪声，以保证模型的训练结果满足全局隐私性。

全局敏感度的计算如公式（16）所示。

$$\Delta f_{global}=max\{\Delta f_i\}$$Δfglobal=max{Δfi}

（16）

为方便对结果进行量化，本文的数据集设置中，各客户端使用的数据量相同。因此，基于全局隐私预算，计算出满足(ε,δ)(ε,δ)差分隐私的高斯噪声大小如公式（17）所示。

$$\sigma_{global}=c\Delta f_{global}\varepsilon_{max}$$σglobal=cΔfglobalεmax

（17）

其中，$c>2ln(1.25/\delta)$c>2ln⁡(1.25/δ)。若当前的全局隐私预算没有达到全局隐私预算阈值ε_{max}εmax 时，中央服务器需要添加的噪声大小如公式（18）所示。

$$\sigma_D=\sigma_{2global}-\bar{\sigma}_{2U}=c\Delta f_{global1\varepsilon_{2max}-1\varepsilon_2}$$σD=σglobal2−σ¯U2=cΔfglobal1ε2max−1ε2

（18）

其中，$\sigma_U = c\Delta f_{global}/\varepsilon_k$σU=cΔfglobal/εk。

通过在联邦学习的客户端和服务器端同时实施差分隐私保护，可以在两个关键层面，即客户端个体信息层面和服务器全局模型层面提供双重隐私保障。在训练前期，对客户端添加较大的噪声，从而有效保护单个用户的数据信息；随着训练的进行，客户端添加的噪声逐渐减小，而服务器端添加的噪声会增大，这种方法能够在维持全局隐私保护水平的同时，减小客户端上传梯度参数的偏移程度，从而达到加速模型收敛的效果。

总之，本节提出的全局隐私调整策略成功平衡了隐私保护和模型性能，不仅能够保证联邦学习参与用户的隐私得到充分保护，还能提高模型的整体安全性。

2.4 算法总览

1）将一种自适应加噪的差分隐私机制应用到联邦学习客户端的局部模型中。具体来说，根据每个客户端特定的隐私需求设定了个性化的隐私预算，并且随着训练过程的推进动态调整隐私预算值，以实现细粒度的隐私保护。

2）在中央服务器端设计并实施了一种模型参数随机化重排策略，进一步增强模型更新中的个体信息的混淆程度。在该策略中，一个可信赖的中央节点负责接收来自各客户端上传的模型参数，并对这些模型参数进行重新排列，对于处于相同层次梯度参数的不同客户端，中央节点进行随机交换操作，从而进一步提高隐私保护的强度。

3）根据预先设定的全局隐私预算阈值ε_{max}εmax，中央服务器端将在聚合的梯度参数中添加适量

的噪声，以满足全局隐私预算的要求。这一步骤旨在确保在整个联邦学习过程中，整体模型的隐私保护水平得到最终调整，并维持一致的隐私保护标准。

通过以上 3 个步骤，既保证了客户端数据的隐私性，又确保了联邦学习模型的有效性和准确性，整体算法流程如算法 3 所示。

算法 3 基于异构数据的联邦学习自适应差分隐私算法

输入：客户端隐私预算$\{\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_m\}$$\{\varepsilon1,\varepsilon2,\cdots,\varepsilon m\}$，全局模型参数和客户端上传梯度$(w_t,g_{ki})$(wt,gik)，梯度裁剪阈值$M$M，全局隐私预算阈值$\varepsilon_{max}$εmax，客户端数$m$，模型参数层数$L$L，参数洗牌器模型每一层交换次数$N$N

输出：扰动后的全局模型参数

1. for k in range(epoch_num):for k in range(epoch_num):

客户端：

2. for j in range(m):for j in range(m):

3. 根据算法敏感度计算高斯噪声分布$N(0,\Delta f_2\sigma_2)$N(0,Δf2σ2)

4. 根据裁剪阈值对局部模型梯度进行裁剪：$g_{kj}=g_{kj}\cdot M/\!/g_{kj}/\!/_2$gjk=gjk⋅M∥gjk∥2

5. 对梯度添加高斯扰动：$g_{kj}\sim\leftarrow g_{kj}+N(0,\Delta f_2\sigma_2)$gjk~←gjk+N(0,Δf2σ2)

参数洗牌器：

6. for t=1 to L:for t=1 to L://参数层数

7. for i=1 to N:for i=1 to N://每一层交换次数

8. 生成$[m]$[m]区间的两个不相等的随机整数：*a*,*b*

9. 客户端 *a* 和客户端 *b* 的第 *t* 层梯度参数进行交换：$swap((a,b),t)$swap((a,b),t)

10. 得到洗牌后的梯度参数：$\{g_{k1'},g_{k2'},\cdots,g_{km'}\}${g1k′,g2k′,⋯,gmk′}

11. 将洗牌后的梯度参数上传到服务器

服务器：

12. 根据全局隐私预算阈值ε_{max}εmax 计算添加噪声大小

$$\sigma_D=\sigma_{2global}-\bar{\sigma}_{2U}----------\surd=c\Delta f_{global1\varepsilon_{max2}-1\varepsilon_2}---------\surd\sigma D=\sigma global2-\bar{\sigma}U2=c\Delta fglobal1\varepsilon max2-1\varepsilon 2$$

13. 梯度聚合：

$\Delta g_{sum}\leftarrow\Sigma_{i=1}^{m}(g_{k'i}+N(c\Delta f_{global1\varepsilon 2max-1\varepsilon 2}---------\surd))$Δgsum←∑i=1m(gik′+N(cΔfglobal1ε2max−1ε2))

14. 计算新一轮全局模型：

$w_{k+1}\leftarrow w_{k+1}\frac{1}{n}\Sigma_{i=1}^{n}(g_{ki}+N(O,\Delta f^2\sigma^2))$wk+1←wk+1n∑i=1n(gik+N(0,Δf2σ2))

15. return w_{epoch_num}wepoch_num

3 实验及分析

本文的实验环境如表 1 所示。

表 1　实验环境及基本参数配置

名称	配置信息
操作系统	Windows 10
CPU	Intel(R) Core(TM) i7-7500U
GPU	NVIDIA RTX 4070(24 GB)
框架	PyTorch 1.10.0+cuda11.1
内存	16 GB
开发语言	Python 3.8

本文实验的一些参数设定：学习率为 0.01，本地批处理大小 batch_size 为 32，本地迭代次数为 1，全局迭代次数 epochs 为 250。

本文在来自 5 个基准数字分类数据集上进行实验，不同数据集具有异构的外观，即每个数据集有属于自身的特征，但是共享相同的标签和标签分布。具体来说，本文实验的数据集为 MNIST、SVHN、USPS、SynthDigits、MNIST-M 这 5 种数字图像分类数据集，为了严格控制非相关因素（如客户端之间的样本数不平衡），将这 5 种数据集分别作为 5 个客户端的私有数据，每个客户端从数据集中随机选择等量的训练样本进行联邦学习训练，其中 80%用作训练集，20%用作测试集，保证每个客户端拥有的数据都具有独立的特征，并且相互之间具有相同的标签和标签分布。

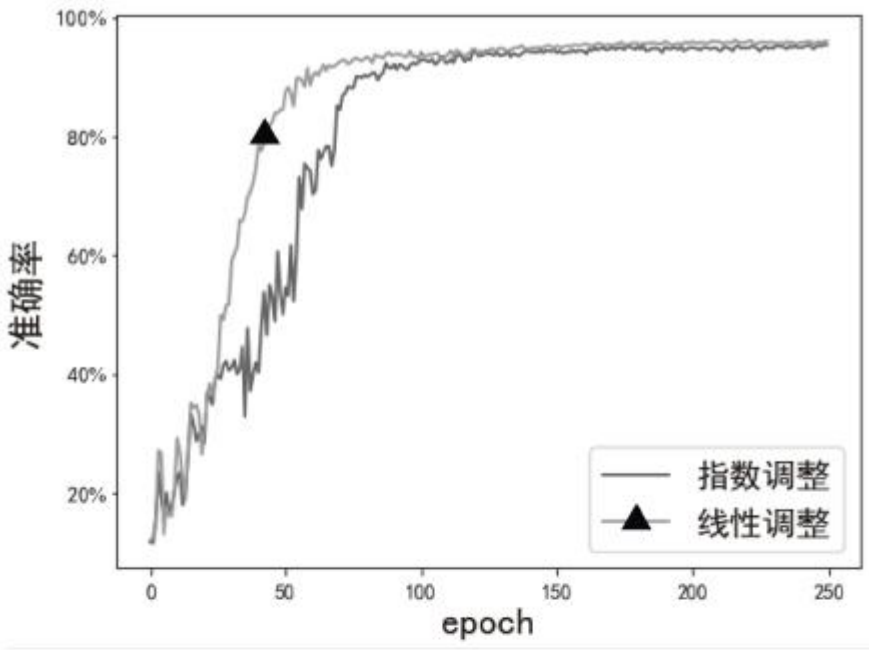
3.1 自适应调整策略对比

为了分析线性调整和指数调整策略两种隐私预算自适应调整方法的适用性，在相同因素按预算设定条件下，分别使用这两种策略进行模型训练，比较收敛过程和模型效果。

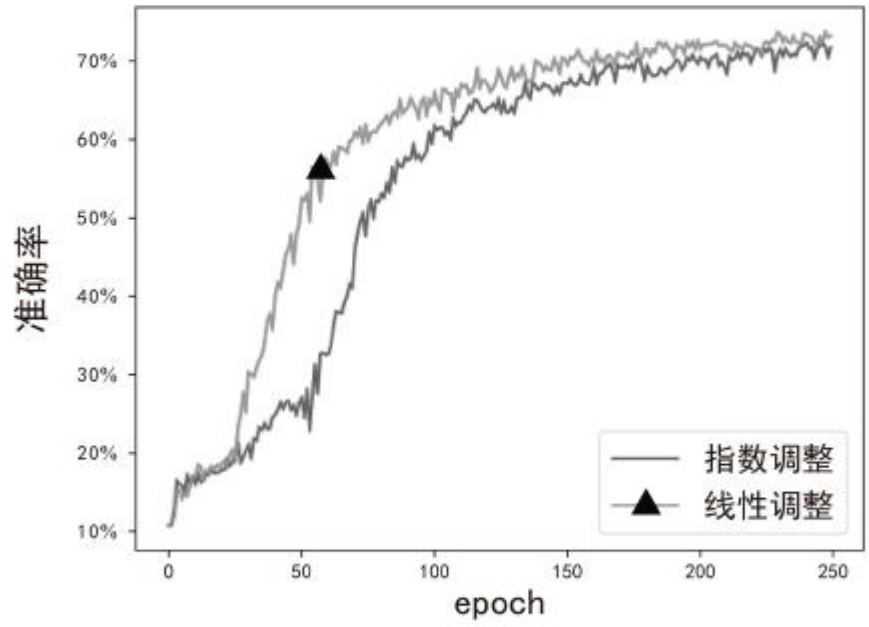
本节实验的隐私预算设定为{MNIST:4、SVHN:4、USPS:8、SynthDigits:8、MNIST-M:12}，隐私阈值为 36。

对比实验的结果如图 6 所示。

图 6



a) MNIST



b) MNIST–M

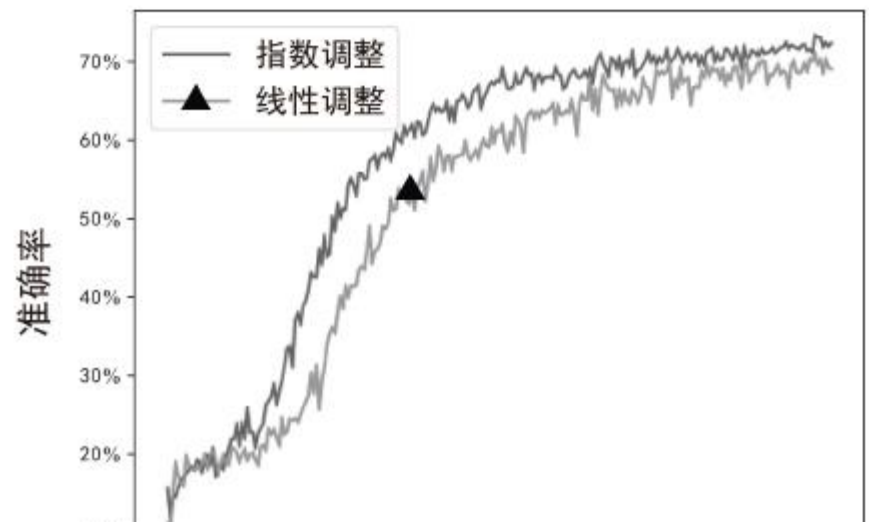


图 6　自适应调整策略对比

根据图 6，在同样的隐私预算设定条件下，在 MNIST、MNIST-M、SynthDigits 和 USPS 客户端上，线性调整策略实验的模型训练收敛速度更快，收敛后的模型效果相似，原因是在模型训练前期，与指数调整相比，线性调整的速率更高，所以各个客户端的隐私预算调整更多，噪声的影响更小，在训练后期，指数调整的速率更高，客户端的隐私预算增加，在总隐私预算一致的情况下，训练得到的模型效果相似；在 SVHN 客户端上，指数调整策略实验的模型训练收敛速度更快，收敛后的模型效果略好，原因是模型收敛的方向有利于该客户端模型的收敛方向，在训练前期，loss 值下降较多，指数调整的幅度更大，导致该客户端隐私预算更大，噪声更小，收敛速度更快。

从总体来看，线性调整策略更有利于加速模型收敛，并且与指数计算相比，线性计算消耗的算力更少，在本节之后的实验中，实验方法均使用线性调整策略。

3.2 隐私阈值ε_{max} ε max 对准确率的影响

在联邦学习框架中，隐私阈值ε_{max}εmax 的设置是为了在中央服务器端调节系统的整体隐私保护与

水平。根据隐私预算调整方法，提高各个客户端的隐私预算，从而降低对客户端梯度参数添加的噪声大小，根据隐私阈值在服务器端添加补偿噪声对全局隐私调整可以保证整个联邦学习系统的隐私保护程度。

实验将 MNIST、SVHN、USPS、SynthDigits、MNIST-M 这 5 个数据集对应的客户端的初始隐私预算分别设定为 1、1、2、2、3，隐私阈值ε_{max}εmax=9。同时，设置了两个消融对比实验，分别是

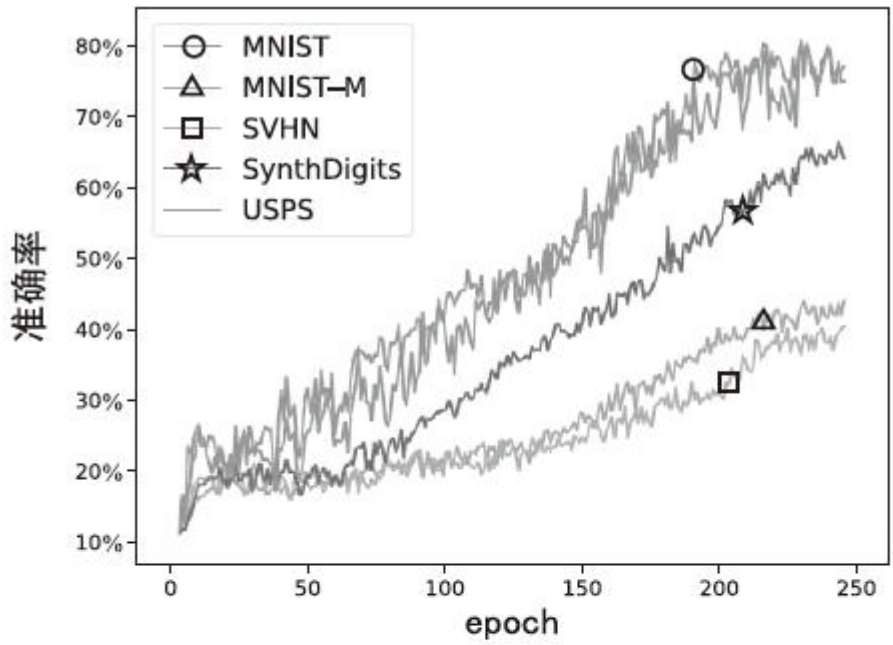
实验 1 和实验 2。

1）实验 1：在同样隐私预算设定下，不设定隐私阈值，隐私预算随着训练进程无限地自适应增加。

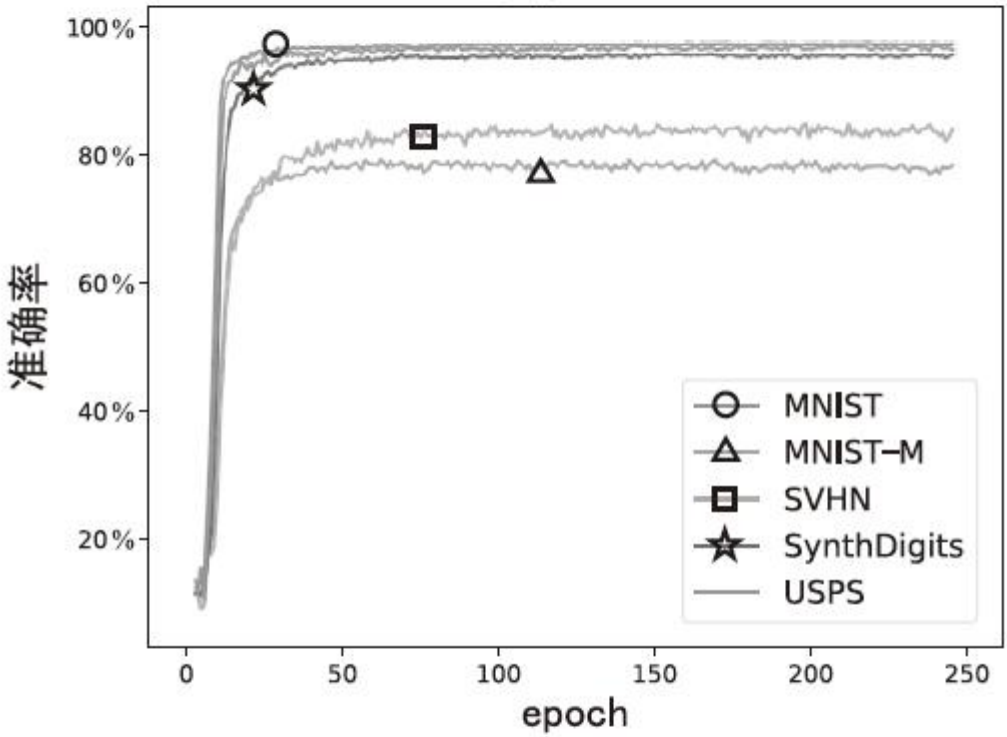
2）实验 2：在同样的隐私预算设定下，训练过程中不对隐私预算进行调整。

根据图 7，实验 1 不设定隐私阈值，随着训练的过程，隐私预算无限地自适应增加，到训练后期，隐私阈值增加到较大值，噪声尺度随之巨幅降低，对模型收敛的影响微乎其微，但是其对模型参数的隐私保护程度也大大降低，无法满足隐私需求；实验 2 训练过程中不调整隐私参数，满足总隐私预算的噪声全部添加在客户端上传的梯度参数上，由于噪声尺度较大，模型在 250 个 epoch 之后仍未收敛；本文方法与实验 2 的总隐私预算大小相同，但是在模型训练过程中对隐私预算执行了自适应调整策略，由于设置的初始隐私预算较低，模型在 250 个 epoch 之后趋于收敛。

图 7


MNIST
MNIST–M
SVHN
SynthDigits
USPS
准确率
epoch
80%
70%
60%
50%
40%
30%
20%
10%
0
50
100
150
200
250

a）本文方法


MNIST
MNIST–M
SVHN
SynthDigits
USPS
准确率
epoch
100%
80%
60%
40%
20%
0
50
100
150
200
250

b）实验 1

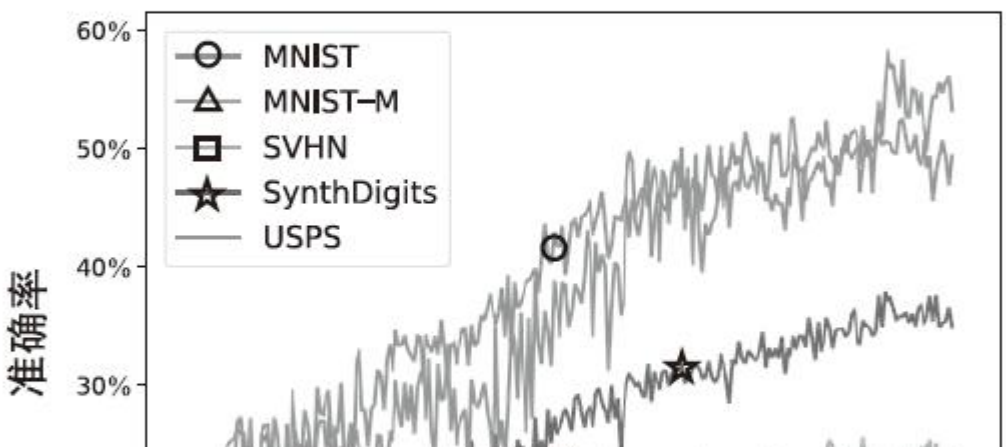
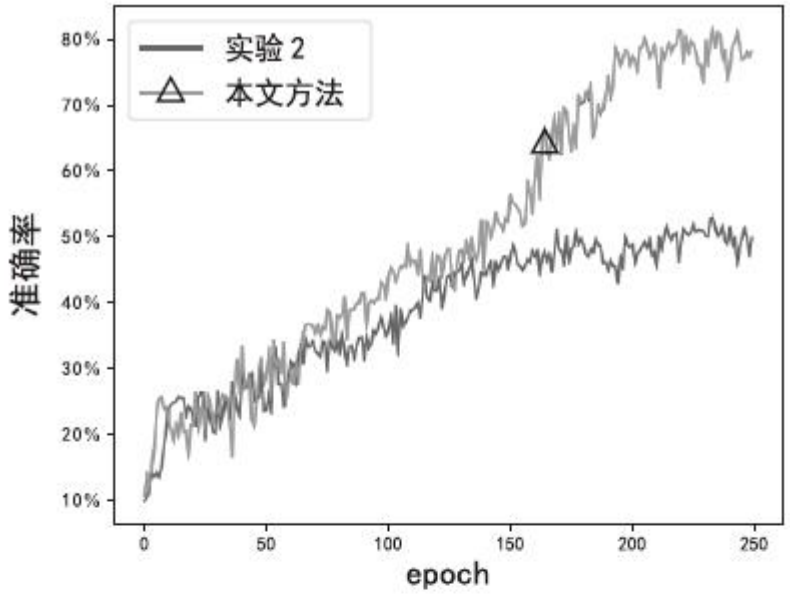

MNIST
MNIST–M
SVHN
SynthDigits
USPS
准确率
60%
50%
40%
30%

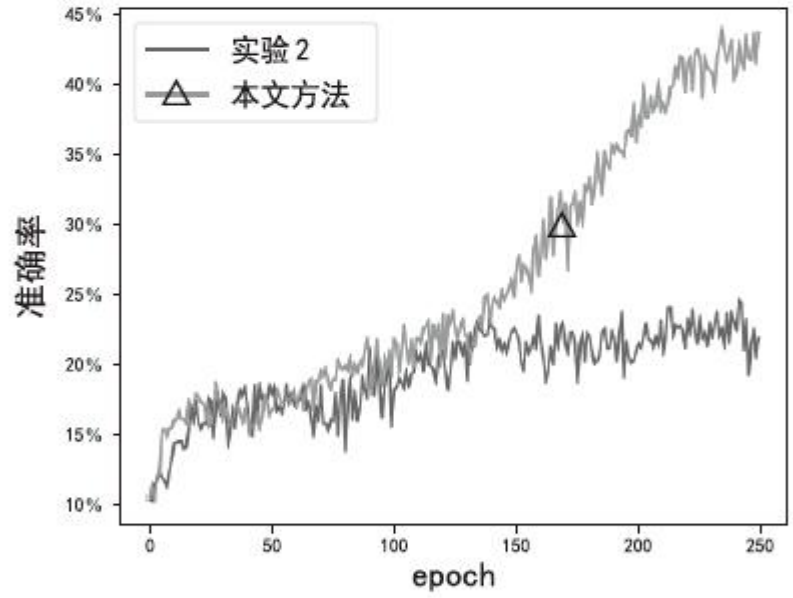
图 7　不同隐私阈值设定对模型收敛的影响

图 8 是实验 2 与本文方法的对比结果，分别对比了 5 个客户端下两种方法的收敛进程和收敛速度。根据结果可以看出，对于任意一个客户端，在模型训练前期，本文方法和实验 2 的模型在测试集上的准确率均维持在较低水平，并且准确率比较接近，在训练后期，本文方法的准确率明显高于实验 2。本文方法在训练后期，随着 loss 值的降低，噪声尺度随之降低，同时，中央服务器对全局模型参数进行隐私修正，使整个系统的隐私预算满足设定值，由于异构数据分布的特性，添加噪声较大会增加客户端训练模型的偏移程度，使全局模型更难以收敛。于是，在同样的总隐私预算条件下，本文方法在训练后期取得的模型准确率与实验 2 相比具有显著提升。

图 8



a) MNIST



b) MNIST–M

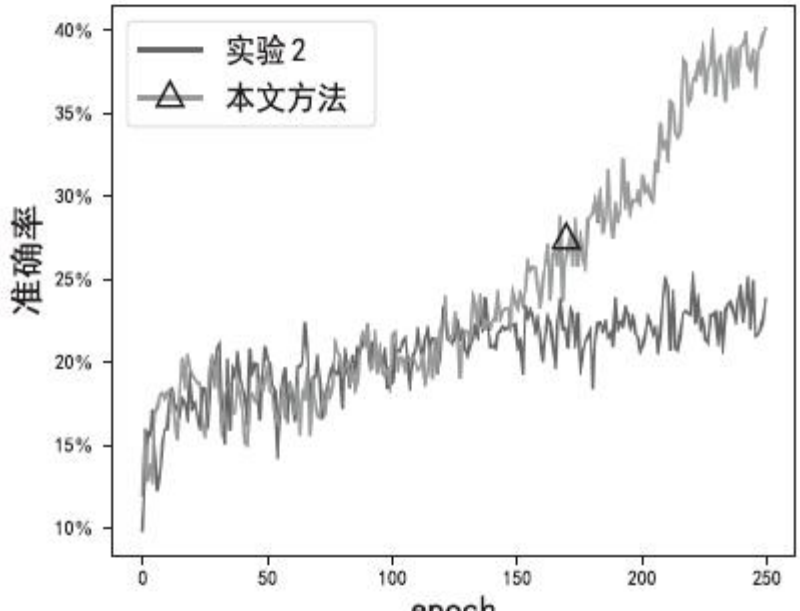


图 8　实验 2 与本文方法对比

3.3 噪声对准确率的影响

基于 4 个隐私预算设定，通过实验结果分析噪声大小对联邦学习模型收敛性能的影响。隐私预算设定如下。

1）设定 1。{MNIST:1、SVHN:1、USPS:2、SynthDigits: 2、MNIST-M:3}，隐私阈值为 9。

2）设定 2。{MNIST:2、SVHN:2、USPS:4、SynthDigits: 4、MNIST-M:6}，隐私阈值为 18。

3）设定 3。{MNIST:4、SVHN:4、USPS:8、SynthDigits: 8、MNIST-M:12}，隐私阈值为 36。

4）设定 4。{MNIST:8、SVHN:8、USPS:16、SynthDigits: 16、MNIST-M:24}，隐私阈值为 72。

隐私预算参数如表 2 所示。

表 2　隐私预算参数

客户端	设定 1	设定 2	设定 3	设定 4
MNIST	1	2	4	8
SVHN	1	2	4	8
USPS	2	4	8	16
SynthDigits	2	4	8	16
MNIST-M	3	6	12	24
总隐私预算	9	18	36	72
隐私阈值	9	18	36	72

通过实验对比，本文将 4 种隐私预算设定条件下的模型训练效果总结在表 3 中。

表 3　噪声对准确率影响对比试验结果

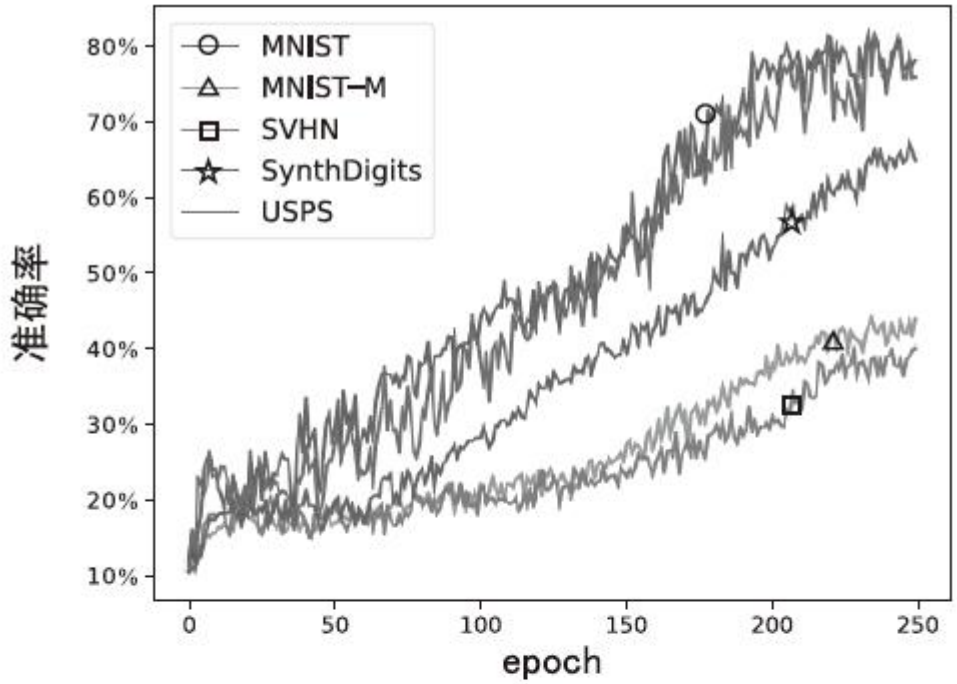
客户端	设定 1	设定 2	设定 3	设定 4

客户端	设定 1	设定 2	设定 3	设定 4
MNIST	79.23%	87.51%	95.47%	97.36%
SVHN	38.63%	55.31%	70.27%	78.98%
USPS	75.81%	86.32%	95.23%	96.64%
SynthDigits	65.72%	80.28%	90.29%	93.61%
MNIST-M	43.78%	59.04%	71.78%	76.54%

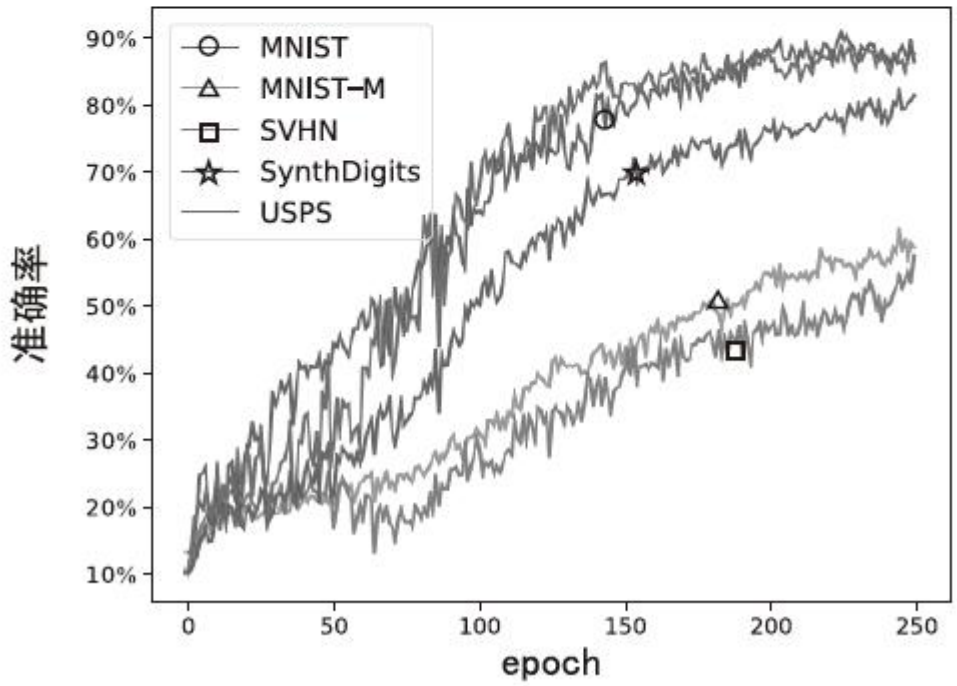
根据结果可知，随着隐私预算的增加，在 5 个客户端下模型分类的准确率随之增加，模型的可用性随之提高，表明本文方法符合差分隐私规律。

为了进一步分析在不同隐私预算设定下的模型收敛过程，在 4 种设定下的对比实验的模型收敛过程如图 9 所示。

图 9


MNIST
MNIST–M
SVHN
SynthDigits
USPS
准确率
80%
70%
60%
50%
40%
30%
20%
10%
0
50
100
150
200
250
epoch

a）设定 1


MNIST
MNIST–M
SVHN
SynthDigits
USPS
准确率
90%
80%
70%
60%
50%
40%
30%
20%
10%
0
50
100
150
200
250
epoch

b）设定 2

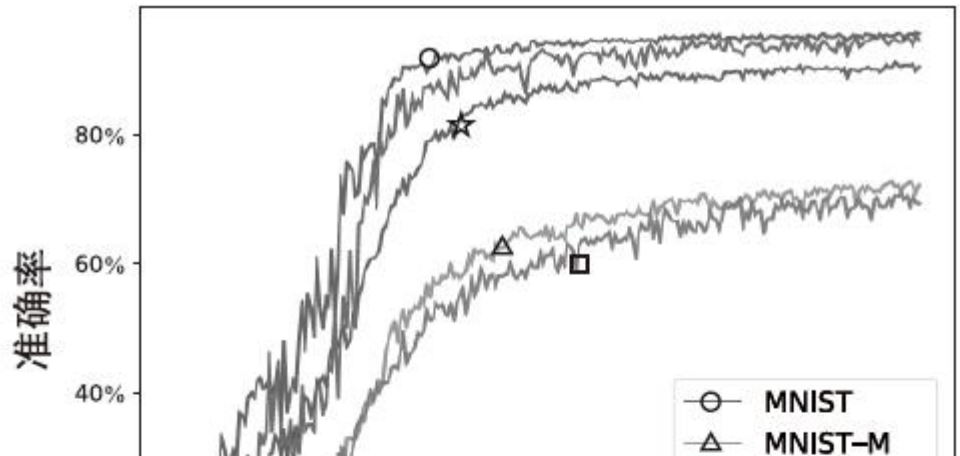

准确率
80%
60%
40%
MNIST
MNIST–M

图 9　模型收敛过程

根据图 9a）和图 9b），在设定 1 和设定 2 的隐私预算下，经过 250 个 epoch 后联邦学习仍然没有收敛，表明较小的隐私预算造成了较大的参数扰动，模型需要进行更多轮次迭代才能达到收敛条件，并且收敛的模型效果较差；根据图 9c）和图 9d），在设定 3 和设定 4 的隐私预算下，分别在 170 多轮次和 150 多轮次后模型达到了收敛状态，表明在本节实验的参数和数据条件下，隐私预算设定比较合理，能够较好地权衡模型性能和隐私保护效果，体现了本节方法的可用性，有一定的参考意义。

3.4 本文算法与其他算法对比

在联邦学习中应用差分隐私机制，向模型参数添加噪声扰动，会影响模型的准确率和收敛速度，并且不同的隐私预算设置、是否使用自适应策略也会导致不同的模型训练效果。同时，在联邦学习模型训练过程中，通信开销是面临的关键问题之一，参与方需要和中央服务器之间进行多轮次交互，直至模型收敛。当模型的参数量和参与训练的设备数量非常庞大时，在交互过程中会产生巨大的通信负担，进而导致多设备参数传输产生延迟，严重影响联邦学习的进程。因此，分析联邦学习中的不同算法通信开销和噪声对模型整体的效率和性能具有重要意义。

本节设置了 3 个实验来分析算法通信开销和噪声对联邦学习模型准确率和收敛速度的影响，实验中的隐私预算设置分别如下。

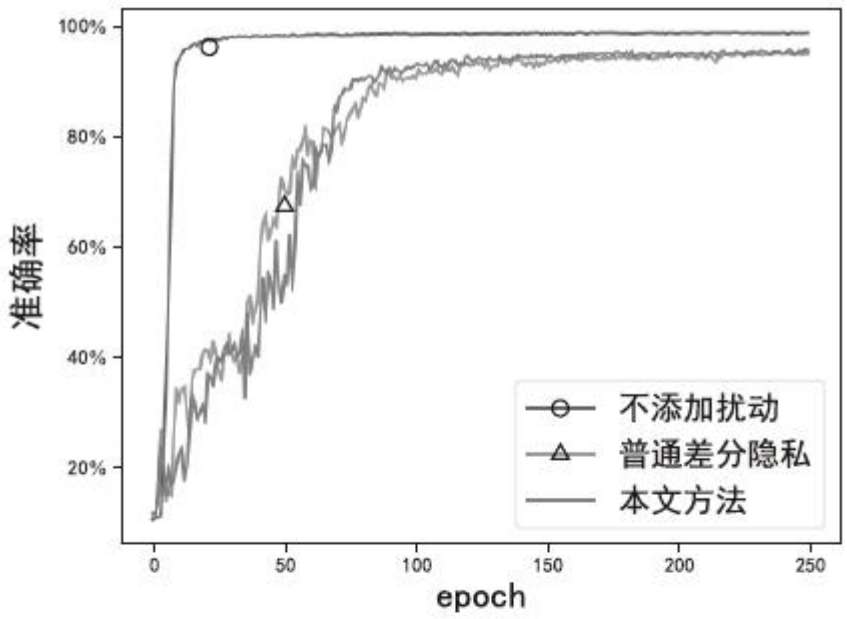
1）实验 3：不应用差分隐私方法。

2）实验 4：应用普通差分隐私方法，预算平均分配为{MNIST:7.2、SVHN:7.2、USPS:7.2、SynthDigits:7.2、MNIST-M:7.2}。

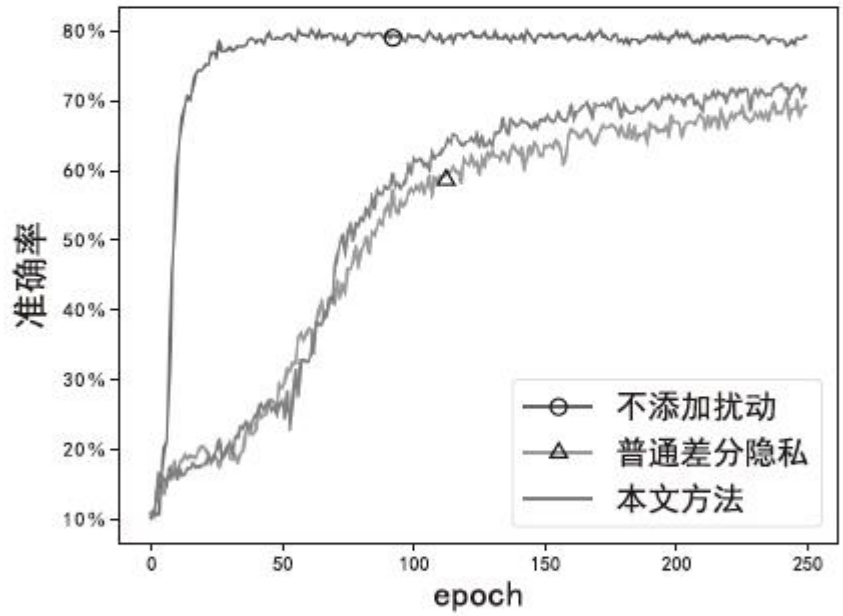
3）实验 5：应用本文提出的自适应差分隐私方法，预算分配为{MNIST:4、SVHN:4、USPS:8、SynthDigits:8、MNIST-M: 12}，隐私阈值为 36。

其中，实验 4 和实验 5 的联邦学习总隐私预算是相等的，对比实验的结果如图 10 所示。

图 10



a）MNIST



b）MNIST–M

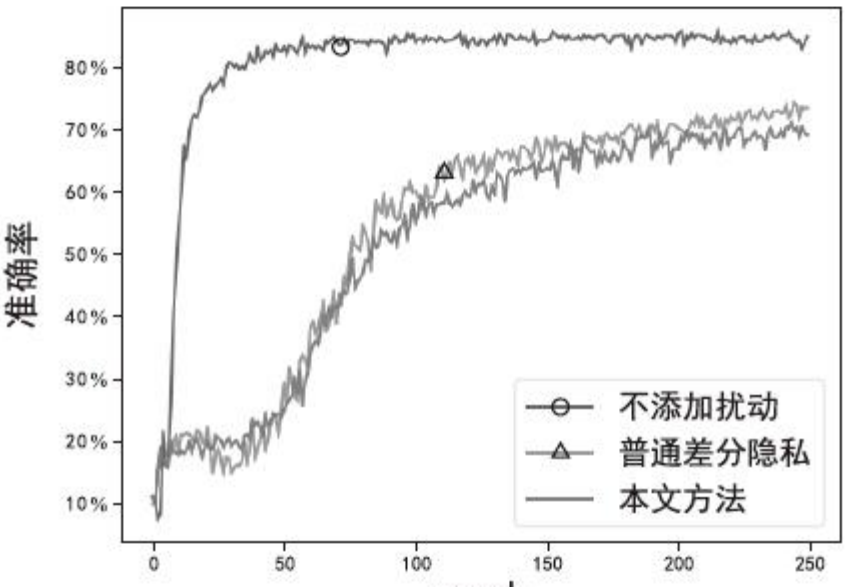


图 10　不同算法分析对比结果

根据图 10，与不应用差分隐私方法对比，普通差分隐私方法和本文提出的自适应差分隐私方法进行联邦学习训练都需要更多的 epoch 来达到收敛，由于客户端和服务器需要更多轮次的交互，增加了模型的通信开销，降低了模型的收敛速度。

同时，根据图 10，噪声显然对于模型训练的收敛速度和模型效果都产生了较大的影响，在相同的总隐私预算条件下，本文提出的自适应调节差分隐私算法与普通差分隐私算法相比，不同客户端的收敛速度受到预设定的隐私预算值的影响，在 MNIST-M 客户端上，本文方法的效果更好，原因是该客户端设定的初始隐私预算值过大，在 SVHN 客户端上，本文方法效果略差，原因是该客户端设定的初始隐私预算值过小，总体看来模型训练收敛之后的整体模型效果与普通差分隐私方法相比效果相似，但是本文方法的个性化差分隐私考虑了不同客户端对于隐私保护的差异性需求，并且参数洗牌机制和中心化差分隐私可以有效避免服务器端的隐私泄露。综合来看，与普通差分隐私方法相比，本文提出的基于异构数据的联邦学习自适应差分隐私方法可以在实现更高程度的隐私保护的同时，不牺牲模型效果，体现了本文方法的优越性。

4 结束语

本文提出一种自适应性差分隐私方法来保护基于异构数据的联邦学习算法。首先，本文在联邦学习的客户端局部模型中采用了自适应加噪的差分隐私方案，针对不同客户端的隐私需求，设定相应的隐私预算，并且在训练的迭代过程中动态调整这些预算值，以适应不同阶段的隐私保护需求；其次，在客户端和服务器之间设置了一个可信的第三方中央节点，负责对客户端上传的模型参数进行重新排列，对不同客户端梯度参数的相同层进行随机交换，可以有效降低恶意服务器推理攻击的可能性，同时中央服务器对梯度参数进行求和平均的操作并没有受影响；最后，在服务器端模型聚合阶段，根据全局隐私预算阈值，在服务端添加适当大小的噪声来调整全局隐私水平，这种全局隐私调整方法可以在保证同等级别隐私保护的同时，实现比传统的本地化差分隐私方法更优的联邦学习模型性能。实验表明，基于本文设定的特征偏移异构数据集，在相等的总预算条件下，本文提出的自适应差分隐私与普通差分隐私相比，能够实现性能更优的联邦学习模型。

参考文献

View Option

[1]

XIONG Shiqiang, HE Daojing, WANG Zhendong, et al.

A Review of Federated Learning and its Security and Privacy Protection

[J]. Computer Engineering, 2024, 50(5): 1-15.

DOI:10.19678/j.issn.1000-3428.0067782 [本文引用: 1]

熊世强, 何道敬, 王振东, 等.

联邦学习及其安全与隐私保护研究综述

[J]. 计算机工程, 2024, 50(5): 1-15.

DOI:10.19678/j.issn.1000-3428.0067782 [本文引用: 1]

[2]

ACAR A, AKSU H, ULUAGAC A S, et al.

A Survey on Homomorphic Encryption Schemes: Theory and Implementation

[J]. ACM Computing Surveys (Csur), 2018, 51(4): 1-35.

[本文引用: 1]

[3]

LI Xiang.

On the Convergence of Fedavg on Non-IID Data

[EB/OL]. (2020-06-25)[2024-09-15]. https://doi.org/10.48550/arXiv.1907.02189.

URL [本文引用: 1]

[4]

ZHAO Yue.

Federated Learning with Non-IID Data

[EB/OL]. (2022-07-21)[2024-09-15]. https://doi.org/10.48550/arXiv.1806.00582.

URL [本文引用: 1]

[5]

ABADI M, CHU A, GOODFELLOW I, et al.

Deep Learning with Differential Privacy

[C]// ACM. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2016: 308-318.

[本文引用: 1]

[6]

DU Wenliang, ATALLAH M J.

Secure Multi-Party Computation Problems and their Applications: A Review and Open Problems

[C]// ACM.Proceedings of the 2001 Workshop on New Security Paradigms. New York: ACM, 2001: 13-22.

[本文引用: 1]

[7]

HASHEMI H, WANG Yongqin, ANNAVARAM M.

DarKnight: An Accelerated Framework for Privacy and Integrity Preserving Deep Learning Using Trusted Hardware

[C]// ACM. MICRO-54: 54th Annual IEEE/ACM International Symposium on Microarchitecture. New York: ACM, 2021: 212-224.

[本文引用: 1]

[8]

MOHAMMADI N, BAI Jianan, FAN Qiang, et al.

Differential Privacy Meets Federated Learning under Communication Constraints

[J]. IEEE Internet of Things Journal, 2021, 9(22): 22204-22219.

[本文引用: 1]

[9]

GONG Xuan, SONG Liangchen, VEDULA R, et al.

Federated Learning with Privacy-Preserving Ensemble Attention Distillation

[J]. IEEE Transactions on Medical Imaging, 2022, 42(7): 2057-2067.

[本文引用: 1]

[10]

GAO Dashan, LIU Yang, HUANG Anbu, et al.

Privacy-Preserving Heterogeneous Federated Transfer Learning

[C]// IEEE. 2019 IEEE International Conference on Big Data (Big Data). New York: IEEE, 2019: 2552-2559.

[本文引用: 1]

[11]

NOBLE M, BELLET A, DIEULEVEUT A.

Differentially Private Federated Learning on Heterogeneous Data

[C]// PMLR. International Conference on Artificial Intelligence and Statistics. New York: PMLR, 2022: 10110-10145.

[本文引用: 1]

[12]

YANG Li, ZHU Lingbo, YU Yueming, et al.

Review of Federal Learning and Offensive-Defensive Confrontation

[J]. Netinfo Security, 2023, 23(12): 69-90.

[本文引用: 1]

杨丽, 朱凌波, 于越明, 等.

联邦学习与攻防对抗综述

[J]. 信息网络安全, 2023, 23(12): 69-90.

[本文引用: 1]

[13]

MAMMEN P M.

Federated Learning: Opportunities and Challenges

[EB/OL]. (2021-01-14)[2024-09-15]. https://doi.org/10.48550/arXiv.2101.05428.

URL [本文引用: 1]

[14]

YANG Liuyan, HE Juanjuan, FU Yue, et al.

Federated Learning for Medical Imaging Segmentation via Dynamic Aggregation on Non-IID Data Silos

[J]. Electronics, 2023, 12(7): 1687-1707.

[本文引用: 1]

[15]

LI Qinbin, DIAO Yiqun, CHEN Quan, et al.

Federated Learning on Non-Iid Data Silos: An Experimental Study

[C]// IEEE. 2022 IEEE 38th International Conference on Data Engineering (ICDE). New York: IEEE, 2022: 965-978.

[本文引用: 1]

[16]

DWORK C.

Differential Privacy

[C]// Springer. International Colloquium on Automata, Languages, and Programming. Heidelberg: Springer, 2006: 1-12.

[本文引用: 1]

[17]

XU Ruzhi, DAI Lipeng, XIA Diya, et al.

Research on Centralized Differential Privacy Algorithm for Federated Learning

[J]. Netinfo Security, 2024, 24(1): 69-79.

[本文引用: 1]

徐茹枝, 戴理朋, 夏迪娅, 等.

基于联邦学习的中心化差分隐私保护算法研究

[J]. 信息网络安全, 2024, 24(1): 69-79.

[本文引用: 1]

[18]

GIRGIS A M, DATA D, DIGGAVI S, et al.

Shuffled Model of Federated Learning: Privacy, Accuracy and Communication Trade-Offs

[J]. IEEE Journal on Selected Areas in Information Theory, 2021, 2(1): 464-478.

[本文引用: 2]

QQiiPi(x,y)Pi(x,y)iijjPj(x,y)≠Pi(x,y)Pj(x,y)≠Pi(x,y)Pi(x,y)Pi(x,y)Pi(y|x)Pi(x)Pi(y|x)Pi(x)Pi(x|y)Pi(y)Pi(x|y)Pi(y)w∗ w∗ w∗ 1w1∗ w∗ 2w2∗ wt1w1twt2w2twt+1wt+1w∗ w∗ w∗ w∗ w∗ 1w1∗ wt+1wt+1w∗ w∗ ε−ε−δδδδε−ε−δδδδlll